

# Post Nonlinear Independent Subspace Analysis<sup>\*</sup>

Zoltán Szabó and Barnabás Póczos and Gábor Szirtes and András Lőrincz

Department of Information Systems, Eötvös Loránd University,  
Pázmány P. sétány 1/C, Budapest H-1117, Hungary  
{szzoli,pbarn}@cs.elte.hu,szirtes@inf.elte.hu,andras.lorincz@elte.hu

**Abstract.** In this paper a generalization of Post Nonlinear Independent Component Analysis (PNL-ICA) to Post Nonlinear Independent Subspace Analysis (PNL-ISA) is presented. In this framework sources to be identified can be multidimensional as well. For this generalization we prove a separability theorem: the ambiguities of this problem are essentially the same as for the linear Independent Subspace Analysis (ISA). By applying this result we derive an algorithm using the mirror structure of the mixing system. Numerical simulations are presented to illustrate the efficiency of the algorithm.

## 1 Introduction

Independent Component Analysis (ICA) has become one of the most popular research line in signal processing. It aims to solve the following (so called ‘cocktail party’) problem: let us assume there are  $D$  1-dimensional, independent hidden sources and we can only observe their linear mixture through  $D$  microphones. The goal is to recover the original sources from the mixed signals. In spite of its simplicity this model has been successfully applied to, for example, feature extraction problems, denoising or biomedical signal processing. For a recent review on ICA see e.g. [1, 2]. The strong assumption on the linearity can be relaxed by assuming component-wise distortion resulting in a post nonlinear extension (PNL-ICA) of the ICA [3]. This direction has recently gained much attention, for a review see [4]. Another generalization of the original ICA problem is called Multidimensional Independent Component Analysis (MICA) [5], or Independent Subspace Analysis (ISA) (in this paper we use the latter abbreviation). In the ISA framework the assumption on independence is relaxed in the sense that it only requires that groups of the sources be independent so the hidden components can now be multidimensional. Several methods have already been proposed to solve this problem [5–16], and it has also been used in EEG analysis [6].

In this paper we show that both generalizations can be treated at the same time (Post Nonlinear Independent Subspace Analysis, PNL-ISA). PNL-ISA problem arises, for example, if the world is observed through a layer of artificial neural

---

<sup>\*</sup> J. Marques de Sá et al. (Eds.): ICANN, Part I, LNCS 4668, pp. 677–686, 2007.  
© Springer-Verlag Berlin Heidelberg 2007. The original publication is available at [http://dx.doi.org/10.1007/978-3-540-74690-4\\_69](http://dx.doi.org/10.1007/978-3-540-74690-4_69).

networks: neurons mix and sum input signals and then pass the sums through non-linearities.

We prove that multidimensional sources can still be recovered even if the observations are component-wise nonlinear distortions of the linear mixtures. To do so, first we introduce the PNL-ISA problem in Section 2. The ambiguities of the problem are analyzed in Section 3, where a theorem about the separability of the components is proven. Applying this theorem we propose an efficient algorithm in Section 4 that solves the PNL-ISA task. In Section 5 the efficiency of the algorithm is illustrated on several problems. Our results are recapped in Section 6.

## 2 The PNL-ISA Model

Let us define the PNL-ISA task. Let us assume the observations are post non-linear mixtures of multidimensional independent sources:

$$\mathbf{x}(t) = \mathbf{f}[\mathbf{A}\mathbf{s}(t)], \quad (1)$$

where  $\mathbf{s}(t)$  is the concatenation of the components  $\mathbf{s}_m(t) \in \mathbb{R}^d$  that is  $\mathbf{s}(t) = [\mathbf{s}_1(t); \dots; \mathbf{s}_M(t)] \in \mathbb{R}^D$ , ( $D = dM$ ). (For simplicity let the dimension of each component be the same  $d$ .) Our assumptions are the following:

1. Source  $\mathbf{s}$  is *d-independent*, that is  $I(\mathbf{s}_1, \dots, \mathbf{s}_M) = 0$ , where  $I$  denotes the mutual information, and  $\mathbf{s}(t)$  is i.i.d in  $t$ .
2. matrix  $\mathbf{A} \in Gl(D)$  (that is it is of size  $D \times D$  and it is invertible) and it is ‘mixing’. By ‘mixing’ we mean the following: decomposing matrix  $\mathbf{A}$  into blocks of size  $d \times d$  ( $\mathbf{A} = [\mathbf{A}_{ij}]_{i,j=1,\dots,M}$ ,  $\mathbf{A}_{ij} \in \mathbb{R}^{d \times d}$ ), then for any index  $i \in \{1, \dots, M\}$  there exist a pair of indices  $j \neq k \in \{1, \dots, M\}$  for which matrices  $\mathbf{A}_{ij}$  and  $\mathbf{A}_{ik}$  are invertible.
3. function  $\mathbf{f} : \mathbb{R}^D \rightarrow \mathbb{R}^D$  is a component-wise transformation that is  $\mathbf{f}(\mathbf{z}) = [f_1(z_1); \dots; f_D(z_D)]$  and  $\mathbf{f}$  is invertible.

The PNL-ISA problem is to estimate the hidden source components  $\mathbf{s}_m$  knowing only the observations  $\mathbf{x}(t)$  ( $t = 1, \dots, T$ ). For  $d = 1$  we get back the PNL-ICA problem, while for choosing  $\mathbf{f}$  as identity the ISA problem is recovered.

## 3 Ambiguities of PNL-ISA

To solve the PNL-ISA problem it is important to see to what extent we can expect to regain the true sources. The ambiguities of the ICA, PNL-ICA and the ISA problems are well known: in ICA, hidden sources can be recovered up to a scalar multiplier and permutation [17]. In addition to these ambiguities there is an additive scalar term in the PNL-ICA problem [18], while in the ISA problem the  $\mathbf{s}_m$  components can be recovered up to the permutation ambiguity between the components and invertible transformations within each subspace [19].

Because of the PNL assumption the hidden sources can be estimated using the mirror structure of the mixing system, that is  $\hat{\mathbf{s}} = \mathbf{W}\mathbf{g}(\mathbf{x})$  ( $\mathbf{W} \in Gl(D)$ ,  $\mathbf{g} : \mathbb{R}^D \rightarrow \mathbb{R}^D$ , where  $\mathbf{g}$  is a component-wise transformation). It has to be shown, however, that the  $d$ -independence of the resulting  $\hat{\mathbf{s}}$  unequivocally means that the true  $\mathbf{s}$  has been found. The following separability theorem shows that indeed this is the case. This statement can be considered as an extension of the results in [20] for the case  $d \geq 1$ . The proof of the theorem will be based on Lemmas 3.4 and 3.5 of [21]. Due to the limited space these lemmas will only be cited. Let  $C^2(V, \mathbb{R})$  and  $C^\omega(V, \mathbb{R})$  denote the  $V$  (open)  $\subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  2-times continuously differentiable and analytic functions, respectively.

**Theorem 1 (PNL-ISA Ambiguities with Locally-Constant Nonzero  $C^2$  Densities).** *Let (i)  $\mathbf{A}, \mathbf{W} \in Gl(D)$ , be mixing matrices; (ii) let  $\mathbf{s}$  be as above, with at most one  $\mathbf{s}_m$  gaussian component, with existing covariance matrix, with somewhere locally constant density function  $p_S \in C^2(\mathbb{R}^D, \mathbb{R})$ , (iii)  $\mathbf{h} : \mathbb{R}^D \rightarrow \mathbb{R}^D$  is a component-wise bijection with coordinate functions in  $C^\omega(\mathbb{R}, \mathbb{R})$ . In this case, if  $\mathbf{e} := [\mathbf{e}_1; \dots; \mathbf{e}_M] = \mathbf{Wh}(\mathbf{As})$  is  $d$ -independent ( $\mathbf{e}_m \in \mathbb{R}^d$ ) with somewhere locally constant density function, then (i)  $\mathbf{h}(\mathbf{x}) = \mathbf{Lx} + \mathbf{p}$ , where  $\mathbf{L} \in Gl(D)$  is a diagonal matrix and  $\mathbf{p} \in \mathbb{R}^D$ , and (ii) components  $\mathbf{e}_m$  ( $m = 1, \dots, M$ ) recover the hidden sources up to permutation and invertible transformations within each subspace (and maybe up to a constant translation).*

*Proof.* To prove the first statement it suffices to show that if  $\mathbf{e} := \mathbf{Wh}(\mathbf{As})$  is  $d$ -independent then derivatives  $h'_{mi}$  are constant for  $\forall(i, m) \in \{1, \dots, d\} \times \{1, \dots, M\}$  where  $\mathbf{h}_m$  is part of  $\mathbf{h}$  that belongs to subspace  $m$  and  $h_{mi}$  is the  $i^{th}$  coordinate function of  $\mathbf{h}_m : \mathbb{R}^d \rightarrow \mathbb{R}^d$ . It directly implies the second part of the statement as if  $\mathbf{Bs} + \mathbf{Wp}$  (where  $\mathbf{B}$  denotes the matrix product  $\mathbf{WLA}$ ) is  $d$ -independent then  $\mathbf{Bs}$  is also  $d$ -independent. In turn, because of the separability properties of the linear ISA,  $\mathbf{B}$  can recover the hidden components up to permutation and invertible transformations within the subspaces (and maybe up to a constant translation within the subspaces).

To prove the first part, let  $p_E$  and  $p_S$  denote the density functions of  $\mathbf{e}$  and  $\mathbf{s}$ , respectively. Based on the transformation rule of the density functions,  $p_E$  can be given as

$$p_E[\mathbf{Wh}(\mathbf{As})] = |\det(\mathbf{W})|^{-1} \left( \prod_{m=1}^M |\det(\mathbf{h}'_m[(\mathbf{As})_m])|^{-1} \right) |\det(\mathbf{A})|^{-1} p_S(\mathbf{s}), \quad (2)$$

where  $(\mathbf{As})_m \in \mathbb{R}^d$  is part of  $\mathbf{As}$  belonging to subspace  $m$ . In addition,  $\mathbf{e}$  is  $d$ -independent implying that  $p_E(\cdot)$  is  $d$ -separated, that is it can be rewritten as

$$p_E = \otimes_{m=1}^M p_m, \quad (3)$$

where functions  $p_m$  are of  $\mathbb{R}^d \rightarrow \mathbb{R}$ . Let us choose point  $\mathbf{s}^0$ , for which  $p_S(\mathbf{s}^0) > 0$ . Then, there exists an open neighborhood  $U \subseteq \mathbb{R}^D$  of this point, where  $p_S|_U > 0$ , and  $p_S|_U \in C^2(U, \mathbb{R})$ . Let us define  $r(\mathbf{s}) := \ln [|\det(\mathbf{W})|^{-1} |\det(\mathbf{A})|^{-1} p_S(\mathbf{s})]$  on

set  $U$ . It can be shown using Eqs. (2)-(3) that the following relation holds

$$r(\mathbf{s}) = \ln \left[ \left( \prod_{m=1}^M |\det(\mathbf{h}'_m[(\mathbf{A}\mathbf{s})_m])| \right) \left( \prod_{m=1}^M p_m([\mathbf{W}\mathbf{h}(\mathbf{A}\mathbf{s})]_m) \right) \right] \quad (4)$$

$$= \sum_{m=1}^M \ln[|\det(\mathbf{h}'_m[(\mathbf{A}\mathbf{s})_m])|] + \xi_m([\mathbf{W}\mathbf{h}(\mathbf{A}\mathbf{s})]_m), \quad (\mathbf{s} \in U) \quad (5)$$

where  $\xi_m := \ln(p_m)$ , locally at  $\mathbf{s}_m^0$ .  $p_S$  is  $d$ -separated, thus function  $r \in C^2(U, \mathbb{R})$  is *linearly  $d$ -separated*. In other words, it can be written as a direct sum, so  $\partial_i \partial_j r \equiv 0$ , where  $\lfloor \frac{i}{d} \rfloor \neq \lfloor \frac{j}{d} \rfloor$  ( $i, j$  correspond to indices in different subspaces). However, since  $p_m$  and therefore  $\xi_m = \ln(p_m)$  is locally constant, this relation holds (without loss of generality on set  $U$ ) when ignoring terms  $\xi_m$ , since their derivatives are 0. Now, following the reasoning of Lemma 3.5 [21] for the  $d$ -separated function  $[\otimes_{m=1}^M \det(\mathbf{h}'_m)](\mathbf{A}\mathbf{s})$ , where  $\mathbf{A}$  is mixing ( $|\cdot|$ s were dropped since functions  $\mathbf{h}_m$  were assumed to be invertible), we can see that each function  $g_m(\mathbf{v}) = \det[\mathbf{h}'_m(\mathbf{v})](\neq 0)$  satisfies a differential equation

$$g_m H_{g_m} - \nabla g_m (\nabla g_m)^* \equiv \mathbf{C}_m g_m^2 \quad (6)$$

on set  $U_m := \mathbf{A}_m(U)$ , where  $\mathbf{A}_m := [\mathbf{A}_{m1}, \dots, \mathbf{A}_{mM}]$ ,  $\mathbf{C}_m \in \mathbb{R}^{d \times d}$ ,  $\nabla$  stands for the gradient,  $H$  is the Hessian and  $*$  denotes transposition. For this reason - similarly to Lemma 3.4 [21] - functions  $g_m$  can be given in the form  $g_m(\mathbf{v}) = e^{\mathbf{v}^* \mathbf{D}_m \mathbf{v} + \mathbf{b}_m^* \mathbf{v} + c_m}$  ( $\mathbf{v} \in U_m$ ) with suitable  $\mathbf{D}_m \in \mathbb{R}^{d \times d}$ ,  $\mathbf{b}_m \in \mathbb{R}^d$ ,  $c_m \in \mathbb{R}$ . Furthermore, as functions  $\mathbf{h}_m$  are assumed to act on each coordinate separately as  $h_{mi}$ ,  $g_m$  can be written as  $g_m = \otimes_{i=1}^d h'_{mi}$ , we arrive at

$$h'_{mi}(t) = \pm e^{d_{mi} t^2 + b_{mi} t + c_{mi}} \quad (d_{mi}, b_{mi}, c_{mi} \in \mathbb{R}) \quad (7)$$

locally, exploiting that  $h'_{mi} \equiv 0$  is not allowed. Based on our assumption on  $h_{mi}$  it holds on all  $\mathbb{R}(\ni t)$ .

Let us introduce the following notation:  $\mathbf{y} = [\mathbf{y}_1; \dots; \mathbf{y}_M] = \mathbf{A}\mathbf{s}$ , where  $\mathbf{y} \in \mathbf{A}(U)$ . Following the above reasoning for the inverse system  $\mathbf{s} = \mathbf{A}^{-1} \mathbf{h}^{-1}(\mathbf{W}^{-1} \mathbf{y})$ ,  $h_{mi}^{-1}$  can be given similar to (7), with other constants. However, if both  $h_{mi}'$  and  $(h_{mi}^{-1})'$  are of exponential type then it follows that  $d_{mi} = 0$  and  $b_{mi} = 0$  [ $\forall (m, i) \in \{1, \dots, M\} \times \{1, \dots, d\}$ ]. Therefore,  $h_{mi}$ s are affine, that is  $h_{mi}(z) = l_{mi}(z) + p_{mi}$  ( $l_{mi}, p_{mi} \in \mathbb{R}$ ), what we wanted to demonstrate.  $\square$

## 4 Algorithm (Sketch)

The theorem proven above implies that by using an appropriate transformation  $\mathbf{g}$  acting on each coordinate separately, the  $d$ -independence of the estimation  $\hat{\mathbf{s}} = \mathbf{W}\mathbf{g}(\mathbf{x})$  solves the PNL-ISA task. The estimation can be done in two steps:

1. Estimate  $\mathbf{g}$ : according to the  $d$ -dependent central limit theorem [22], term  $\mathbf{A}\mathbf{s}$  can be considered as an approximately gaussian variable, so  $\mathbf{g}$  can be approximated as a 'gaussianization' transformation (see [23, 24] for  $d = 1$ ).
2. Estimate  $\mathbf{W}$ : apply a linear ISA method on the result of the 'gaussianization' transformation.

## 5 Illustrations

Now we illustrate the efficiency of the algorithm presented in Section 4. Test databases are described in Section 5.1. To evaluate the solutions we use a performance measure given in Section 5.2. The numerical results are summarized in Section 5.3.

### 5.1 Databases

To test our PNL-ISA method, we have created 4 datasets ( $\mathbf{s}$ ) shown in Fig. 1.

In dataset *3D-geom* (see Fig. 1(a)) the components  $\mathbf{s}_m$  were random variables uniformly distributed over 3-dimensional geometrical shapes ( $d = 3$ ). There were 6 hidden components ( $M = 6$ ), so the total dimension of the hidden source  $\mathbf{s}$  was  $D = 18$ . In dataset *celebrities* the components  $\mathbf{s}_m$  were generated using cartoons of celebrities.<sup>1</sup> The cartoons were interpreted as density functions and the 2-dimensional coordinate pairs ( $d = 2$ ) were sampled with frequency proportional to the pixel intensity at the given coordinates. We used 10 hidden sources ( $M = 10$ ), resulting in  $D = 20$  (see Fig. 1(c)). The dataset  $A\omega$  is scalable, the number of components ( $M$ ) can be set within a quite broad interval. Here, components  $\mathbf{s}_m$  are random variables of uniform distribution over the letters of the English and the Greek alphabets ( $M \leq 26 + 24 = 50$ ,  $d = 2$ ), see Fig. 1(b). While in the datasets defined so far variable  $\mathbf{s}$  was i.i.d in accord with the PNL-ISA model, in the dataset *IFS* this constraint has been relaxed.<sup>2</sup> Here, components  $\mathbf{s}_m$  are realizations of IFS based 2-dimensional ( $d = 2$ ) self-similar structures. For all  $m$  we have chosen the following triple:  $(\{\mathbf{h}_k\}_{k=1,\dots,K}, \mathbf{p} = (p_1, \dots, p_K), \mathbf{v}_1)$ , where (i)  $\mathbf{h}_k : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  are affine transformations in the form  $\mathbf{h}_k(\mathbf{z}) = \mathbf{C}_k\mathbf{z} + \mathbf{d}_k$  ( $\mathbf{C}_k \in \mathbb{R}^{2 \times 2}$ ,  $\mathbf{d}_k \in \mathbb{R}^2$ ), (ii)  $\mathbf{p}$  is a distribution over the indices  $\{1, \dots, K\}$  ( $\sum_{k=1}^K p_k = 1, p_k \geq 0$ ), and (iii) for the initial value we chose  $\mathbf{v}_1 := (\frac{1}{2}, \frac{1}{2})$ . We generated  $T$  samples in the following way: (i)  $\mathbf{v}_1$  is given ( $t = 1$ ), (ii) an index  $k(t) \in \{1, \dots, K\}$  was drawn according to the distribution  $\mathbf{p}$  and the next sample is generated as  $\mathbf{v}_{t+1} := \mathbf{h}_{k(t)}(\mathbf{v}_t)$ . The resulting series  $\{\mathbf{v}_1, \dots, \mathbf{v}_T\}$  was taken as a hidden source component  $\mathbf{s}_m$  and this way we generated 9 components ( $M = 9$ ,  $D = 18$ ) to make the *IFS* dataset (see Fig. 1(d)).

### 5.2 Performance Measure

Let  $E[\cdot]$  denote the expectation. In the sense of Theorem 1, in ideal case  $\mathbf{s} \in \mathbb{R}^D \mapsto \hat{\mathbf{s}} = \mathbf{W}\mathbf{g}[\mathbf{f}(\mathbf{A}\mathbf{s})] \in \mathbb{R}^D$  is an affine transformation residing within the subspaces. For this reason, the linear transformation  $\mathbf{G}$  that optimally approximates the relation  $\mathbf{s} - E[\mathbf{s}] \mapsto \hat{\mathbf{s}} - E[\hat{\mathbf{s}}]$ , resides also within the subspaces and so it is a *block-permutation matrix*. This block-permutation structure can be measured by the normalized version of the Amari-error [25] adapted to the ISA task [21]. Let us decompose matrix  $\mathbf{G} \in \mathbb{R}^{D \times D}$  into blocks of size  $d \times d$ :  $\mathbf{G} = [\mathbf{G}_{ij}]_{i,j=1,\dots,M}$ .

<sup>1</sup> <http://www.smileyworld.com>

<sup>2</sup> IFS stands for Iterated Function System.

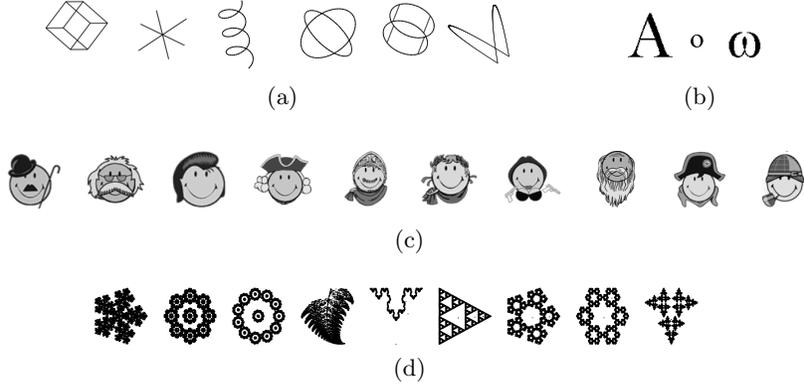


Fig. 1: Illustration of the test datasets. (a): *3D-geom* set. The hidden components  $\mathbf{s}_m$  are random variables of uniform distribution over 3-dimensional geometric shapes:  $d = 3$ ,  $M = 6$ ,  $D = 18$ . (b):  $A\omega$  set. Here, components  $\mathbf{s}_m$  are 2-dimensional variables of uniform distribution over the shape of the letters of the English and Greek alphabets:  $d = 2$ ,  $M \leq 50$ . (c): *celebrities* dataset. The components are 2-dimensional variables with sampling frequency proportional to the pixel intensity of the corresponding coordinates of ‘celebrities’ cartoons:  $d = 2$ ,  $M = 10$ ,  $D = 20$ . (d): *IFS* dataset. Here components  $\mathbf{s}_m$  are not i.i.d. variables anymore, instead they are self-similar structures generated from iterated function systems:  $d = 2$ ,  $M = 9$ ,  $D = 18$ .

Let  $g_{ij}$  denote the sum of the absolute values of matrix  $\mathbf{G}_{ij} \in \mathbb{R}^{d \times d}$ . Now, the following term

$$r(\mathbf{G}) := \frac{1}{2M(M-1)} \left[ \sum_{i=1}^M \left( \frac{\sum_{j=1}^M g_{ij}}{\max_j g_{ij}} - 1 \right) + \sum_{j=1}^M \left( \frac{\sum_{i=1}^M g_{ij}}{\max_i g_{ij}} - 1 \right) \right] \quad (8)$$

denotes the Amari-index that takes values in  $[0,1]$ : for an ideal block-permutation matrix  $\mathbf{G}$  it takes 0; for the worst case it takes 1.

### 5.3 Simulations

In this section the results of our numerical simulations on the above defined datasets are presented. We focused on the following questions:

1. The error of the source estimation as a function of (i) the sample size, (ii) the dimension ( $D$ ) of the task.
2. Estimation of  $D$  when it is not known beforehand.

The gaussianization (see Section 4) was based on the ranks of samples [23], in the solution of the ISA task we followed the method described in [14]. The goodness of the estimation was measured by the Amari-index introduced in

Section 5.2. For a given sample size  $T$  the goodness of 50 random runs ( $\mathbf{A}, \mathbf{s}, \mathbf{f}$ ) were averaged.  $\mathbf{A}$  was a random orthogonal matrix. The nonlinear functions  $f_i$  have been generated as

$$f_i(z) = c_i[a_i z + \tanh(b_i z)] + d_i, \quad (9)$$

that is they are mixtures of random, scaled and translated  $id$  and  $\tanh$  functions. Here  $a_i \in [0, 0.5]$ ,  $b_i \in [0, 5]$ ,  $d_i \in [0, 2]$  are random variables of uniform distribution,  $c_i$  take  $\pm 1$  values with probability  $\frac{1}{2}, \frac{1}{2}$ .

First we studied the Amari-index as a function of the sample size. For  $3D\text{-geom}$ ,  $celebrities$  and  $IFS$  the dimension and the number of the components  $d, M$  and the dimension of the task  $D$  were fixed while for  $A\omega$   $M$  has been chosen as 2, 3, 4, 10, 20, 50. As for small  $M$  the gaussian assumption on  $\mathbf{A}\mathbf{s}$  is less likely, we expect to see deterioration of the goodness of the estimation in this range. In all cases the sample size  $T$  was chosen between 1,000 and 100,000. The Amari-index as a function of the sample size is shown in Fig. 2. For  $T = 100,000$ , the exact errors are shown in Table 1 and Table 2. As an example, the estimation results for  $IFS$  and  $3D\text{-geom}$  datasets are illustrated in Fig. 3.

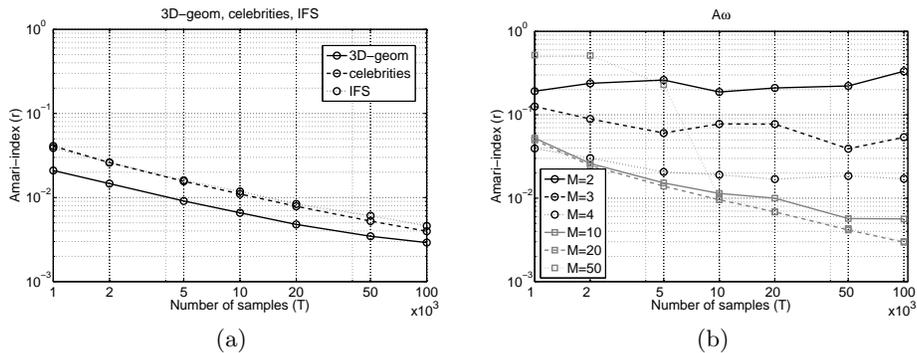


Fig. 2: Average Amari-index as a function of the sample size, on loglog scale. (a): datasets  $3D\text{-geom}$ ,  $celebrities$  and  $IFS$ . (b):  $A\omega$  with different number of components ( $M$ ). For  $T = 100,000$ , the exact errors are shown in Table 1 and Table 2.

| 3D-geom              | celebrities          | IFS                  |
|----------------------|----------------------|----------------------|
| 0.29% ( $\pm 0.05$ ) | 0.40% ( $\pm 0.03$ ) | 0.46% ( $\pm 0.06$ ) |

Table 1: Amari-index for  $3D\text{-geom}$ ,  $celebrities$  and  $IFS$  datasets: average  $\pm$  std. Sample size:  $T = 100,000$ . The error as a function of sample size  $T$  is plotted in Fig. 2(a).

As Fig. 2(a) shows, the dependency of the Amari-index on the sample size  $T$  follows a power-law for datasets  $3D\text{-geom}$ ,  $celebrities$  and  $IFS$ ,  $r(T) \propto T^{-c}$  ( $c > 0$ ). This can be seen as a linear decrease on the loglog scale. The slope of

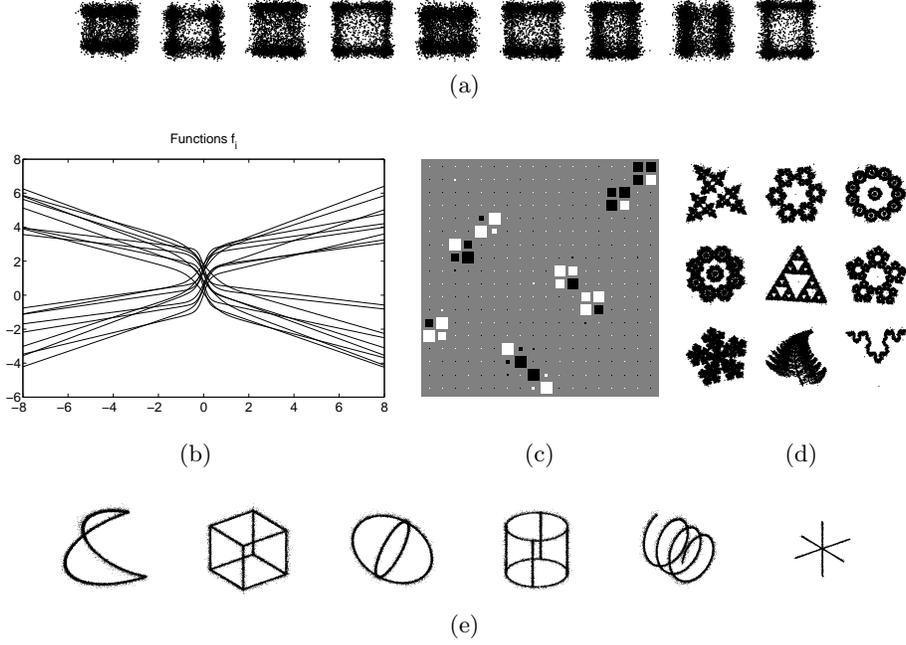


Fig. 3: Illustration of the PNL-ISA estimation on the dataset *IFS* and *3D-geom*, in (a)-(d) and (e), respectively. Sample size:  $T = 100,000$ . (a): the observed mixed  $\mathbf{x}$  signal. (b): the nonlinear  $f_i$  functions. (c) the Hinton-diagram of  $\mathbf{G}$ , ideally it is a block-permutation matrix with blocks of size  $2 \times 2$ . (d): the estimated hidden components ( $\hat{\mathbf{s}}_m$ ). (e): the same as (d), but for the *3D-geom* dataset.

| $M = 2$                | $M = 3$              | $M = 4$              | $M = 10$             | $M = 20$             | $M = 50$             |
|------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| 33.20% ( $\pm 39.42$ ) | 5.37% ( $\pm 8.82$ ) | 1.71% ( $\pm 0.52$ ) | 0.56% ( $\pm 0.50$ ) | 0.30% ( $\pm 0.03$ ) | 0.30% ( $\pm 0.01$ ) |

Table 2: Amari-index for dataset  $A\omega$ , as a function of the number of components  $M$ : values shown are average  $\pm$  std. Sample size:  $T = 100,000$ . The error as a function of sample size  $T$  is plotted in Fig. 2(b).

the lines are about the same for the different datasets. Table 1 summarizes the average estimation errors for sample size 100,000.

In Fig. 2(b) similar power law relation can be found for dataset  $A\omega$  as well, with increasing  $D$ . For  $M \geq 3$ , (that is when  $D \geq 6$ ) the PNL-ISA solution is already efficient. It can also be seen that for the case of  $M = 50$  number of components (that is the dimension of the task is  $D = 100$ ) we need at least 10,000 samples to get a more reliable estimation, while for  $3 \leq M < 50$  2,000 – 5,000 samples are sufficient.

Next we studied to what extent we can guess the overall dimension  $D$  of the hidden source  $\mathbf{s}$  when it is not given beforehand. The dimension of the observation  $\mathbf{x}$  was set to  $D_x = 2D$  ( $D$  of course is not available for the algorithm). The

mixing matrix  $\mathbf{A} \in \mathbb{R}^{D_x \times D}$  was generated by first creating a random orthogonal matrix of size  $D_x \times D_x$  then choosing its first  $D$  columns. Gaussianization has been done on the observations and then we studied the eigenvalues of the covariance matrix of the resulting transformed signal  $\mathbf{g}(\mathbf{x})$ : ideally there are  $D$  positive and  $D_x - D$  (almost) 0 values. We show the ordered eigenvalues on dataset *celebrities* averaged over 50 runs in Fig. 4. It can be seen that exactly half of the eigenvalues are near 0, then there is a big leap. (For datasets *3D-geom* *IFS* we have got similar results, data is not shown.) Figure 4(b) shows the results corresponding to dataset  $A\omega$  for different number of components  $M$ : only the  $M \leq 4$  cases are illustrated, but in the whole range of  $2 \leq M \leq 50$  there is a sharp transition similar to the results gained for the other 3 datasets.

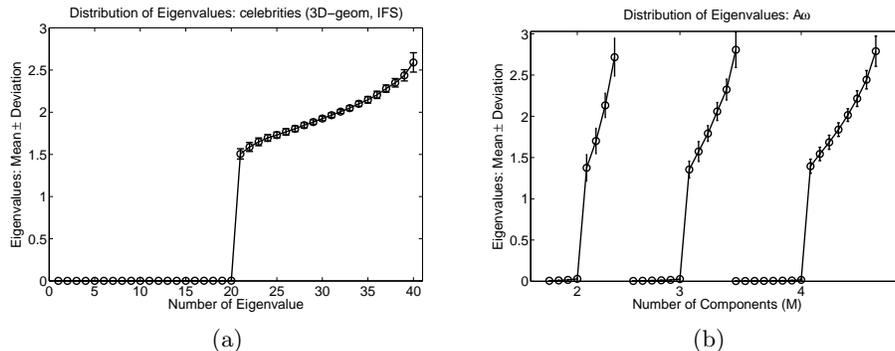


Fig. 4: Estimation of the dimension of the hidden source  $\mathbf{s}$ . The ordered eigenvalues of the covariance matrix of the transformed signal  $\mathbf{g}(\mathbf{x})$  are plotted. Results are averaged over 50 runs. (a): dataset *celebrities*; results for *3D-geom* and *IFS* are similar. (b): eigenvalues for the dataset  $A\omega$ , at different number of components  $M$ .

## 6 Conclusions

In this paper we introduced the PNL-ISA problem as a common extension of the Post Nonlinear Independent Component Analysis (PNL-ICA) and the Independent Subspace Analysis (ISA). We have shown the ambiguities of the PNL-ISA task are essentially the same as in the linear ISA task (up to a constant translation in each subspace). We derived an algorithm based on our separability results. We also demonstrated the efficiency of the algorithm on different datasets. Our simulations revealed that the error of the estimation of the hidden sources decreases in a power law fashion as the sample size increases. This tendency is even more characteristic when the overall dimension of the task ( $D$ ) increases. Interestingly, our algorithm can recover the sources in cases when the assumptions of the PNL-ISA problem are violated: even non i.i.d self-similar hidden

components can be recovered. In addition, we demonstrated that the dimension of the hidden source can also be estimated.

## References

1. Cichocki, A., Amari, S.: Adaptive blind signal and image processing. John Wiley & Sons (2002)
2. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley & Sons (2001)
3. Taleb, A., Jutten, C.: Source separation in post-nonlinear mixtures. *IEEE Transactions on Signal Processing* **10**(47) (1999) 2807–2820
4. Jutten, C., Karhunen, J.: Advances in blind source separation (BSS) and independent component analysis (ICA) for nonlinear systems. *International Journal of Neural Systems* **14**(5) (2004) 267–292
5. Cardoso, J.: Multidimensional independent component analysis. In: Proc. of ICASSP 1998. (Volume 4.) 1941–1944
6. Akaho, S., Kiuchi, Y., Umeyama, S.: MICA: Multimodal independent component analysis. In: Proc. of IJCNN 1999. (Volume 2.) 927–932
7. Hyvärinen, A., Hoyer, P.O.: Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Computation* **12** (2000) 1705–1720
8. Hyvärinen, A., Köster, U.: FastISA: A fast fixed-point algorithm for independent subspace analysis. In: Proc. of ESANN. (2006) 371–376
9. Vollgraf, R., Obermayer, K.: Multi-dimensional ICA to separate correlated sources. In: Proc. of NIPS 2001. (Volume 14.) 993–1000
10. Bach, F.R., Jordan, M.I.: Beyond independent components: Trees and clusters. *Journal of Machine Learning Research* **4** (2003) 1205–1233
11. Póczos, B., Lőrincz, A.: Independent subspace analysis using k-nearest neighborhood distances. In: Proc. of ICANN 2005. (Volume 3697.) 163–168
12. Póczos, B., Lőrincz, A.: Independent subspace analysis using geodesic spanning trees. In: Proc. of ICML 2005. (Volume 119.) 673–680
13. Theis, F.J.: Blind signal separation into groups of dependent signals using joint block diagonalization. In: Proc. of ISCAS 2005. (Volume 6.) 5878–5881
14. Szabó, Z., Lőrincz, A.: Real and complex independent subspace analysis by generalized variance. In: Proc. of ICARN. (2006) 85–88
15. Nolte, G., Meinecke, F.C., Ziehe, A., Müller, K.R.: Identifying interactions in mixed and noisy complex systems. *Physical Review E* **73**(051913) (2006)
16. Theis, F.J.: Towards a general independent subspace analysis. In: Proc. of NIPS. Volume 19. (2006)
17. Comon, P.: Independent component analysis, a new concept? *Signal Processing* **36** (1994) 287–314
18. Achard, S., Jutten, C.: Identifiability of post nonlinear mixtures. *IEEE Signal Processing Letters* **12**(5) (2005) 423–426
19. Theis, F.J.: Uniqueness of complex and multidimensional independent component analysis. *Signal Processing* **84**(5) (2004) 951–956
20. Theis, F.J.: A new concept for separability problems in source separation. *Neural Computation* **16** (2004) 1827–1850
21. Theis, F.J.: Multidimensional independent component analysis using characteristic functions. In: Proc. of EUSIPCO. (2005)

22. Petrov, V.: Central limit theorem for m-dependent variables. In: Proc. of the All-Union Conf. on Probability Theory and Mathematical Statistics. (1958) 38–44
23. Ziehe, A., Kawanabe, M., Harmeling, S., Müller, K.R.: Blind separation of post-nonlinear mixtures using linearizing transformations and temporal decorrelation. *Journal of Machine Learning Research* **4**(7-8) (2004) 1319–1338
24. Solé-Casals, J., Jutten, C., Pham, D.: Fast approximation of nonlinearities for improving inversion algorithms of PNL mixtures and wiener systems. *Signal Processing* **85** (2005) 1780–1786
25. Amari, S., Cichocki, A., Yang, H.H.: A new learning algorithm for blind signal separation. *Advances in Neural Information Processing Systems* **8** (1996) 757–763